Crackers!

Consumer magazine 2002 gives some nutritional information from an analysis of these 18 brands of cracker. Some of this information is shown in the table below:

Energy (Calories/100g)	Number of crackers per100g	Salt (mg/100g)	Total Fat (%)	Saturated Fat (% of all fat)
EE	16	600	2.0	28
385	10	400	2.5	24
408	17	200	3.5	17
405	56	500	4.0	45
411	13	200	4.5	41
405	61	600	5.0	49
413	5	700	7.0	63
419	9	500	7.0	13
426	33	700	8.0	46
429	7	900	9.5	47
451	11	400	14.5	48
484	24	1300	20.5	48
487	23	900	22.5	46
505	21	800	24.0	42
512	16	700	25.0	41
520	61	1000	27.5	43
510	31	1200	28.5	17
536	16	800	30.5	47

1) Write a question that can be answered using bivariate techniques.
2) Plot the data you will use to answer your question and describe, in detail, what is shows.
3) Answer the question asked, using your plot and any numbers obtained from it.
4) Make some predictions using your plot and/or numbers obtained from it. State how reliable you think your predictions are.
5) Select a second pair of variables and carry out a similar analysis. Explain how this new pair either helps (or does not help) answer your first question.
6) State any assumptions made in your analysis, and problems with it and any improvements you can suggest.

Crackers: Model Answers

*Note: this is only a **suggested** set of answers, and other alternatives are acceptable, especially the choice of variables to compare.* Orange text indicates details marker is looking for.

Question:

Is there a relationship between the fat content of the crackers and their energy content;

and can I use that relationship to estimate the energy content from the fat content?

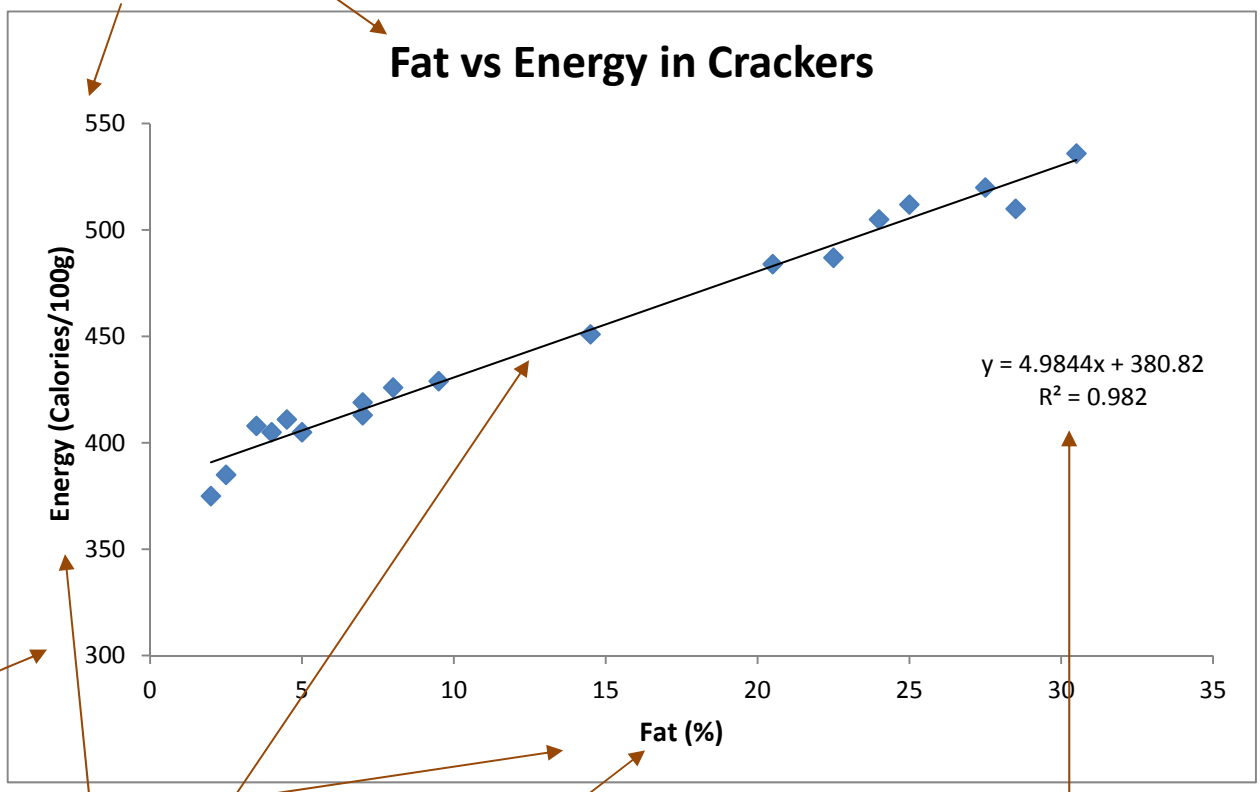
Must relate **two** variables in some sort of **relationship**. Ideally has more than a yes or no answer. (Try to chose the initial variables so that there **is** a relationship, it makes things easier.)

Data used:

The data I used was the fat content vs the energy content columns. I did not omit any data.

If you leave a data point out, for any reason, you must state why.

Properly labelled



Units stated

Line of best fit

Formula and R^2 given

Explanatory variable is on this axis
(Fat content in this case, as it explains energy, not the other way round.)

Plot large enough, with amended starting points on the axes, so that details of the points plotted are clearly visible.

Details of the Plot

The plot shows that there is a strong positive relationship between the fat content of the crackers and their energy content. Clearly state relationship.

The *r* value is 0.99, which is an extremely high correlation. There is almost no chance this is a statistical fluke. Give numbers.

The scatter is even along the length of the line of best fit, with no clear pattern. There were no outlying values. Discuss scatter.

The crackers appear to group into two distinct groups: just over half with the fat below 10% and most of the rest with fat above 20%. One cracker only falls between 10 and 20%. Note features.

Answer to my Question

There is a very strong relationship between the fat content of the crackers and their energy content. Clearly answer the question asked.

A person searching for a low calorie cracker can be confident that a low fat content will mean a low energy content (and *vice versa*).

The R^2 value is 0.982, which means 98% of the variation in the energy content is explained by the fat content. Use R^2 to put a number to the explanatory power of your relationship.

Note that 98% of the energy is not explained, only 98% of the **variation** in the energy content. In this case even a zero fat cracker will have 380 calories.

While this is not proof that the fat is directly resulting in extra calories, as some confounding variable might be working to cause both, this result is consistent with each one per cent increase in fat causing 5 extra calories per 100g of cracker (taken from the slope of the line of best fit). Be careful not to directly attribute cause (without experimental proof) but only state what the results are consistent with.

One potential confounding feature is that the low energy crackers might be “diet” crackers, and thus both low fat and low calories. Discuss why a direct explanation of cause is not possible.

Make Predictions

The formula for the line of best fit is: $y = 4.9844x + 380.82$

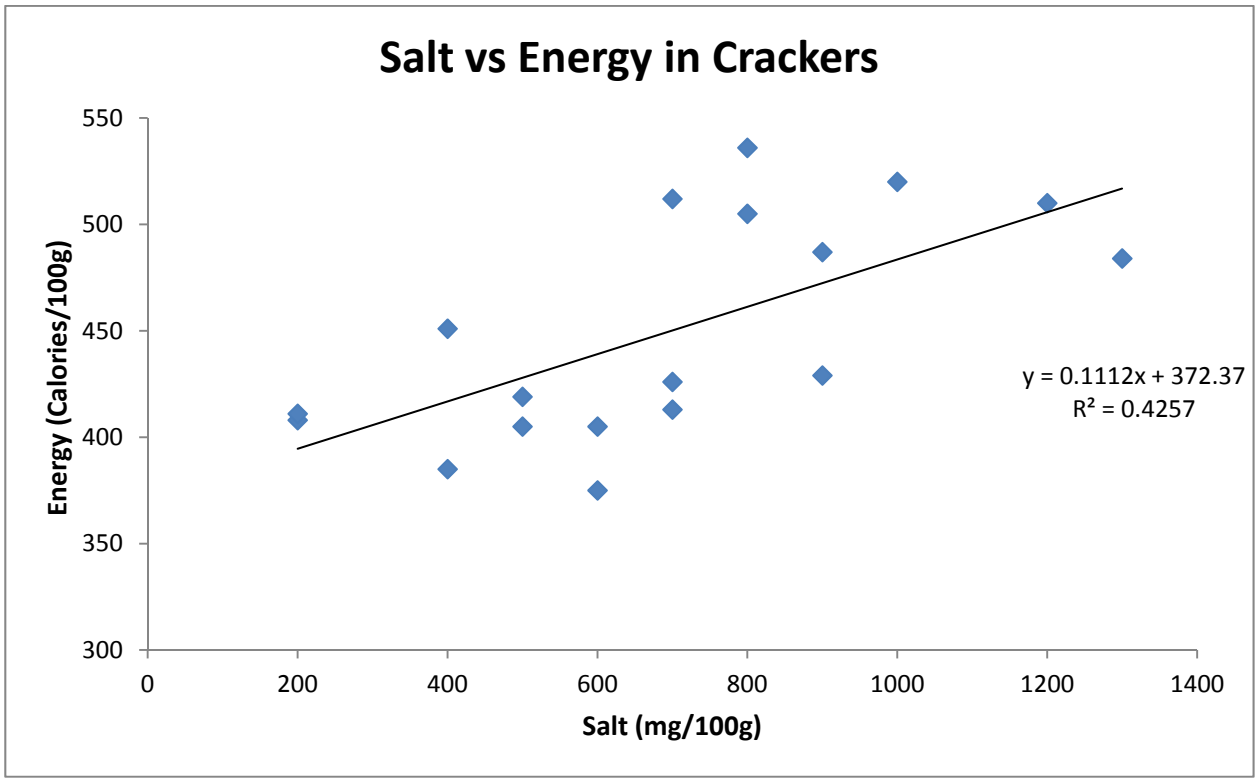
So a fat content of zero will give a cracker with 380.82 calories/100g. Use the formula by preference, although reading the graph will also work.

A person looking for a cracker with under 400 calories/100g should seek one with a fat content under 3.8% (from (400 – 380.82) ÷ 4.9844). Show working so it can be checked.

I think these predictions are extremely accurate, based on the extremely high *r* value of 0.99. Use *r* to explain why a prediction is good or bad.

Second Pair of Variables Merit

I then looked to see if the salt content could explain the energy content to any degree.



There is only a moderate positive relationship between salt and energy, with an *r* value of 0.65.

While the R^2 value of 0.4257 predicted that more than 40% of the variation in energy content could be predicted from the salt content, it seems likely that there is some lurking variable causing this, since salt has no calorie content. Still want numbers.

Perhaps manufacturers of high fat crackers are also seeking a market that doesn't care much about salt content either, while dieters are also likely to be looking at keeping their salt content down. Speculation about lurking variables is OK, but don't push it too far. Just make a reasonable suggestion, then stop.

There is a some grouping in the data, with one group being those around 400 calories/100g and another around 500, and a wide gap in the middle between them. If these are actually two entirely different sets of crackers (say diet and normal) then there is almost no correlation between salt and energy for the two groups. This would need to be thoroughly examined before making a firm conclusion. Discuss features of the data, don't just look at *r* and R^2.

Conclusion

The variation in a cracker's calorie content can be quite adequately predicted using its fat content, roughly calories = 380 + 5 × % of fat.

The salt content is a guide to the calorie content, but only in general terms, and only if there are not two distinct calorie groups in the data.

Suggestions and Improvements Excellence

This analysis assumes all the crackers are the same basic type.

The grouping of the data would need to be examined before more firm conclusions can be drawn. Perhaps instead of diet/non-diet there might be a division between style of cracker – say wheat cracker and maize cracker.

If the data is clumped, then each group should be analysed separately. That is unlikely to make much difference to the link between fat and energy, but may destroy any relationship between salt and energy.

More data might help, but in the fat/energy case only really for the extreme cases of very low fat and very high fat.

It is possible that the saturated fat content could be relevant. An analysis of that might show a more exact relationship for fat to energy (although there isn't much room to improve with *r* already at 0.99)

There seems to be no evidence of any non-linear or piecewise patterns.